Structure, Vol. 11, 1499–1511, December, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.str.2003.11.005

The Structure of Chorismate Synthase Reveals a Novel Flavin Binding Site Fundamental to a Unique Chemical Reaction

John Maclean,[1,*] and Sohail Ali[2]
[1]Department of Structural Biology
[2]Department of Biology
PanTherix Ltd.
Todd Campus
West of Scotland Science Park
Glasgow G20 0XP
United Kingdom

Summary

The crystal structure of chorismate synthase (CS) from *Streptococcus pneumoniae* has been solved to 2.0 Å resolution in the presence of flavin mononucleotide (FMN) and the substrate 5-enolpyruvyl-3-shikimate phosphate (EPSP). CS catalyses the final step of the shikimate pathway and is a potential therapeutic target for the rational design of novel antibacterials, antifungals, antiprotozoals, and herbicides. CS is a tetramer with the monomer possessing a novel β-α-β fold. The interactions between the enzyme, cofactor, and substrate reveal the structural reasons underlying the unique catalytic mechanism and identify the amino acids involved. This structure provides the essential initial information necessary for the generation of novel anti-infective compounds by a structure-guided medicinal chemistry approach.

Introduction

Chorismate synthase (CS) catalyses the seventh and final step of the shikimate biosynthetic pathway, which leads to the biosynthesis of all aromatic compounds. This pathway is found in bacteria, fungi, plants, and apicomplexan parasites, but is absent from mammals. It has long been recognized as a drug target for the generation of broad-spectrum antibacterial and antifungal drugs (Haslam, 1993), and the recent identification of the shikimate pathway in apicomplexan parasites (Roberts et al., 1998) has established further therapeutic possibilities. Several of the enzymes of the shikimate pathway catalyze reactions that are mechanistically interesting, and crystal structures of the first six enzymes have provided significant mechanistic insights (Shumilin et al., 1999, Carpenter et al., 1998, Gourley et al., 1999, Michel et al., 2003, Krell et al., 1998, Stallings et al., 1991). CS catalyses the most interesting and unusual reaction of the entire pathway, the conversion of EPSP to chorismic acid (chorismate), via the 1,4-antielimination of phosphate and a proton, a reaction that is unique in nature (Bornemann, 2002, Macheroux et al., 1999).

CS is particularly attractive as an anti-infective target as chorismate lies at a metabolic node, being the precursor for five distinct pathways. It is necessary for the production of aromatic amino acids, para-aminobenzoic acid (PABA), folate, and for other cyclic metabolites such as ubiquinone and menaquinone. Much of the folate pathway is also absent in mammals, and enzymes within it have therefore been successfully exploited as targets for antibacterial chemotherapy, as exemplified in the inhibition of dihydrofolate reductase (DHFR) by trimethoprim. Inhibition of CS would lead directly to a drop in intracellular concentration of PABA, the substrate for the folate biosynthesis enzyme dihydropteroate synthase (DHPS). This enzyme is the site of action of the well-known sulphonamide series of antibacterial compounds. There is thus scope for a dual therapy approach involving a CS inhibitor and an antifolate.

CS activity has an absolute requirement for reduced flavin mononucleotide (FMN), and although catalytic turnover results in no overall change in redox state, there is considerable evidence that different FMN species are present over the course of the reaction (Macheroux et al., 1996a). A semiquinone form of FMN has been proposed to be a short-lived flavin intermediate in a reaction that proceeds by a free radical mechanism (Osborne et al., 2000). Several nonradical mechanisms have also been advanced, each based on a different short-lived flavin intermediate, which could be formed during the conversion of EPSP to chorismate. The high-resolution structure of CS, which we present herein, permits the validity of these proposed mechanisms to be evaluated.

The structure of CS from *Streptococcus pneumoniae* has been determined to 2.0 Å resolution. The structure is of the ternary complex of CS, cocrystallized with oxidized FMN and EPSP. The structure is analogous to the complex formed by the active enzyme but CS is unable to turn over because of its absolute requirement for reduced FMN. The structure demonstrates that the unique reaction mechanism is dependent on a novel β-α-β architecture and a novel mode of flavin binding. Elucidation of the interactions between CS, FMN, and EPSP provide a detailed understanding of ways in which the active site could be exploited in the design of novel anti-infective compounds.

Results and Discussion

Structure of Chorismate Synthase

The crystal structure of SpCS was solved at 2.0 Å using the multiwavelength anomalous dispersion (MAD) method from a single crystal of selenomethionine-incorporated protein. Details of X-ray data collection and refinement of the final model are given in Tables 1 and 2.

The crystal structure of SpCS reveals a tetrameric molecule showing approximate 222 symmetry. Accessible surface calculations using SURFACE (CCP4) have confirmed that the basic structural unit of CS is a dimer, and that two such dimers compose the observed tetramer. Around 25% of the surface of each monomer is buried upon formation of the dimer, while an additional 12.5% of the surface of each monomer is buried on forming the tetramer. One of the four monomers present

*Correspondence: john.maclean@btinternet.com

Table 1. Crystallographic Data

Crystallographic Data				
	High Resolution	Peak	Inflection Point	Remote
Wavelength (Å)	0.9788	0.9755	0.9790	0.8855
Diffraction limit (Å)	1.9	2.5	2.5	2.8
Completeness (%)	99.9 (99.6)	99.6 (99.3)	99.5 (98.5)	99.6 (99.5)
No. unique reflections	202809	105430	106407	75744
Redundancy	1.9	1.9	1.9	1.8
I/σ_I	13.8 (2.2)	17.0 (2.8)	16.0 (2.5)	25.4 (5.5)
Rmerge (%)	5.7 (33.4)	4.9 (30.8)	5.2 (38.0)	3.7 (15.5)

in the asymmetric unit differs from the other three in the conformations of two loops close to the active site. This makes the active site considerably more accessible, and this monomer is therefore described as having an "open" conformation, while the others are "closed." Differences between open and closed forms provide vital insights into the mobility of these active site loops, which play an important mechanistic role, and are discussed in more detail below. It is clear that differences in the crystal contacts made by the monomers are responsible for the observation of the open form of CS. An adjacent symmetry-related molecule makes close contacts with the open monomer and occupies some of the space into which the mobile active site loops would move in order to close the active site. The open form of CS must therefore represent a conformation that is accessible to the protein when in solution.

The overall fold of the SpCS monomer has been shown to be novel by comparison with known structures using the programs CE (Shindyalov and Bourne, 1998) and VAST (Madej et al., 1995). Comparison with the entire contents of the protein databank (PDB) (Berman et al., 2000) using CE produced only eight hits above the threshold for weak structural similarity, all of which were well below the threshold for membership of existing structural families. Superimposing these hits onto the structure of CS demonstrated that there was no meaningful structural or topological similarity. A similar comparison using VAST also failed to detect significant similarity between CS and existing structural families represented in the PDB.

CS has been predicted to be an α-β barrel as a result of secondary structure prediction efforts (Macheroux et al., 1998, White et al., 1994), but the monomer structure, as shown in Figure 1A, is clearly very different from what was expected. It comprises a single large core domain, which is surrounded by loops and discrete stretches of secondary structure. The core consists of an internal layer of four long α helices, sandwiched between a pair of four-stranded antiparallel β sheets. Such β-α-β secondary structure arrangements are very uncommon and only a few are described in the CATH database of structural domain classifications (Orengo et al., 1997). It is clear from the topology of CS (Figure 2A) that the protein core has pseudo 2-fold symmetry, although the molecule as a whole does not. The pseudo 2-fold axis divides the CS core into a pair of 4β-2α fragments, the secondary structure elements of which can be superimposed with an RMSD of just 1.5 Å. Despite the obvious structural similarity there is no conservation of sequence between these fragments, and hence the origin of the architecture of CS is an intriguing puzzle.

Table 2. Refinement Statistics

Refinement Statistics	
Resolution (Å)	2.0
Final R factor (25–2.0 Å)	15.69
R_{free} (10% total data)	22.24
Rmsd bond lengths (Å)	0.013
Rmsd bond angles (°)	2.585
Number of protein residues	1552
Number of water molecules	1925
Ramachandran analysis (%) (favored/allowed/generous/disallowed)	92.8/7.2/0.0/0.0
Average B factor protein (Å^2)	26.2
Average B factor solvent (Å^2)	43.1
Average B factor FMN	21.1
Average B factor EPSP	31.9

The 4β-2α core fragment is sufficiently simple for there to be topological similarity between it and existing protein structures, the most notable of which are the histidine-containing phosphocarrier protein HPr (PDB code: 1PTF) and the ribosomal recycling factor (PDB: 1EK8). HPr is a small protein with a β sandwich structure, while RRF contains a small α-β domain. There is topological similarity between these small domains and the core fragment, but this does not appear to be significant in comparison with the overall CS structure. In addition, both HPr and RRF lack some of the secondary structure elements present in the CS fragment that are essential to the overall β-α-β architecture.

As shown in Figure 2A, β sheet 1 contains the N terminus of the protein, and consists of strands β1, β2, β7, and β4 in an antiparallel arrangement. β sheet 2 is also antiparallel, and comprises strands β8, β9, β13, and β10. The central helical layer is primarily hydrophobic and comprises helices α1, α4, α8, and α7, arranged up-down-down-up.

The major feature of the CS dimer interface is the extension of β sheet 2 from each monomer into an eight-stranded antiparallel β sheet (Figure 1B). The two sheets come together at strand β10, forming four hydrogen bonds, but there are many other interactions at the dimer interface. The other core secondary structure element involved in stabilization of the dimer is helix α7, which is adjacent to strand β10 in the CS monomer, and buries a considerable amount of hydrophobic surface at the dimer interface. Several other regions of the structure are involved in dimerization, notably helices α2 and α3, loops L9, L10, L15, L16, L17, and L19, and strands β11 and β12, which extend from the monomer core to pack

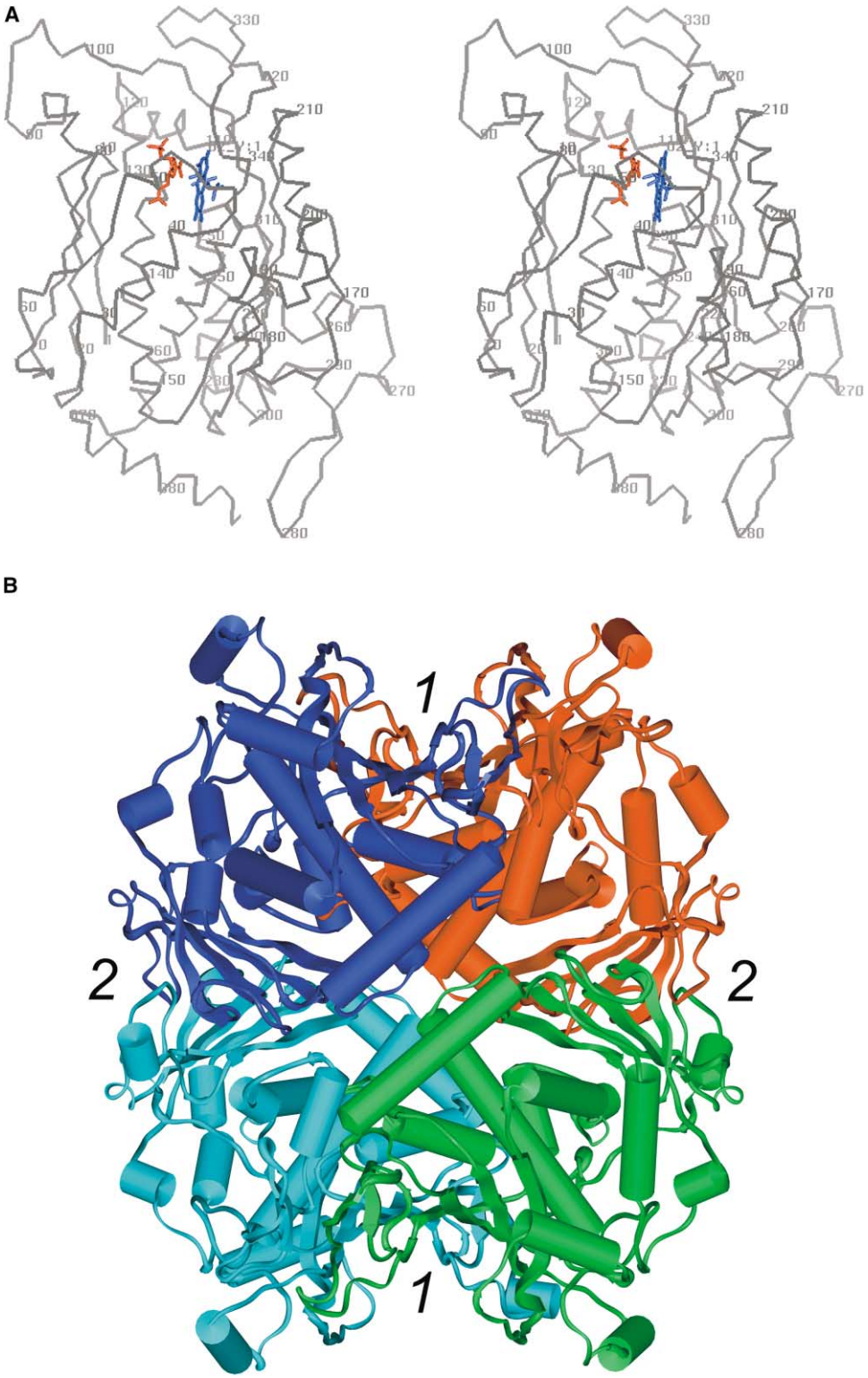
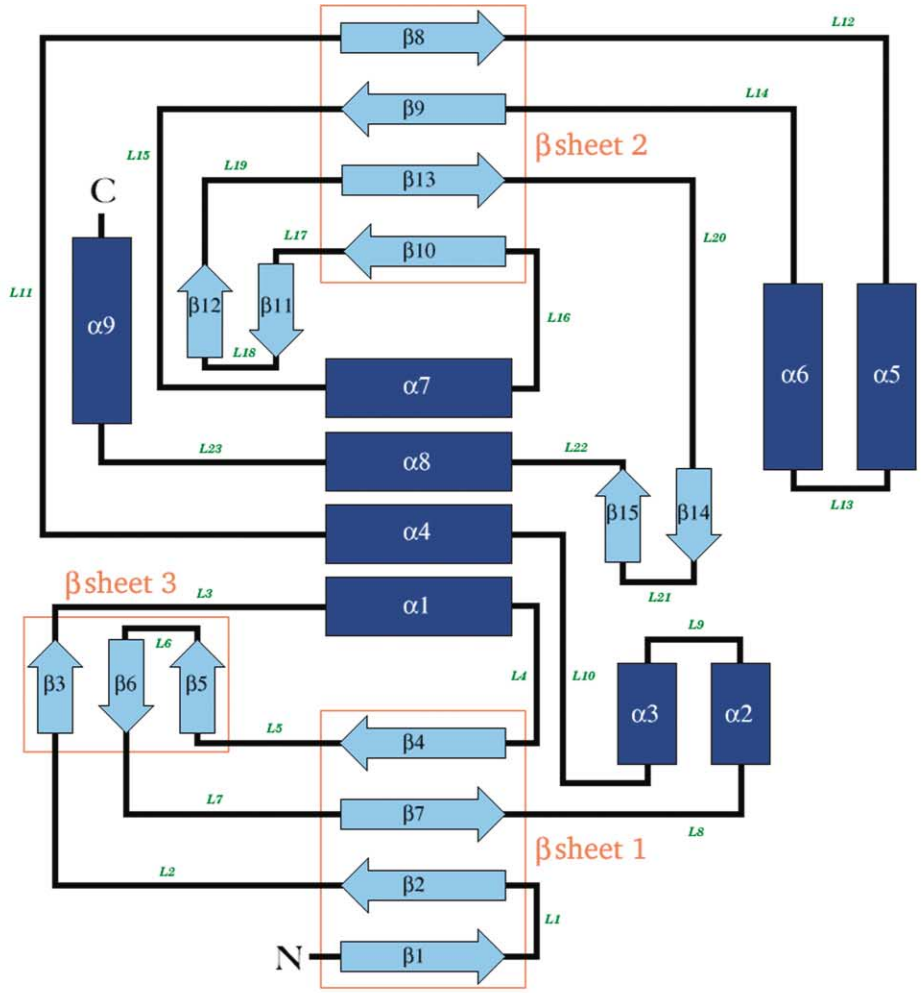
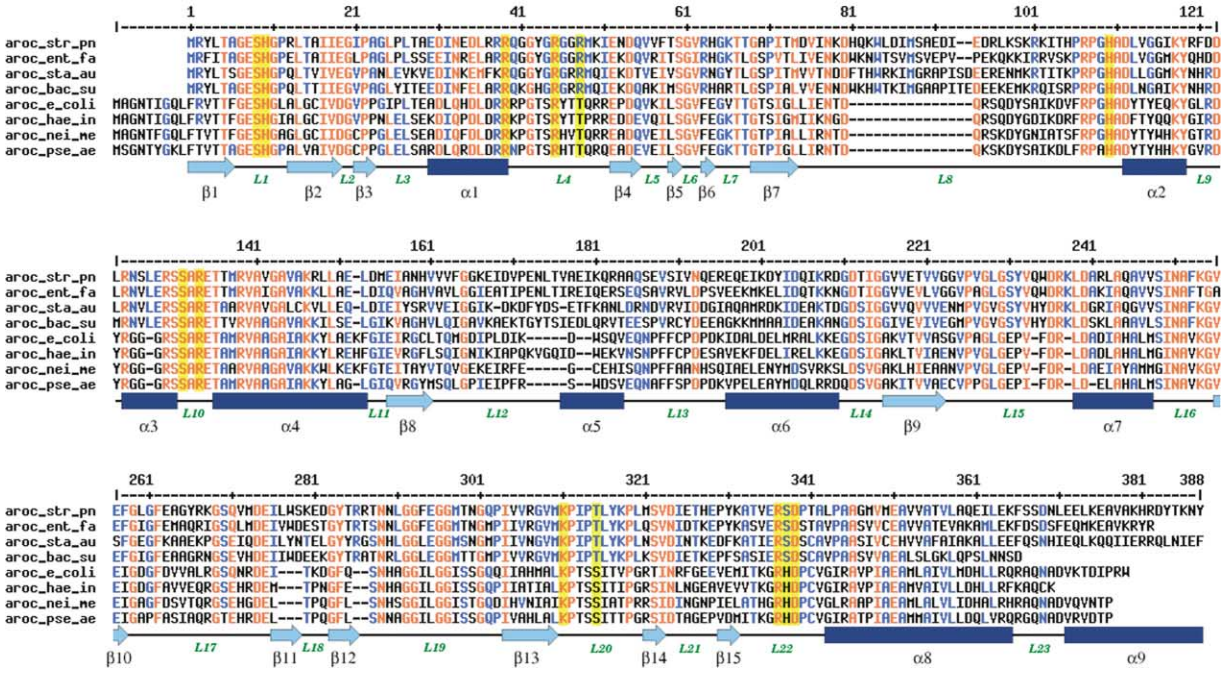


Figure 1. Overall Structure of Chorismate Synthase

(A) A stereo Cα trace of the CS monomer, with every tenth residue marked. FMN is highlighted in blue, EPSP in red.

(B) Ribbon representation of the SpCS tetramer. The β sheet that bridges two monomers, and is a fundamental component of the monomer-monomer interaction that forms the dimer, is denoted *1*. The β sandwich interaction between dimers is denoted *2*.


β8
β9
β13
β10
β12
β11
α9
α7
α8
α4
α1
β3
β6
β5
β4
β7
β2
β1
α6
α5
β15
β14
α3
α2
C
N
β sheet 2
β sheet 3
β sheet 1
L1
L2
L3
L4
L5
L6
L7
L8
L9
L10
L11
L12
L13
L14
L15
L16
L17
L18
L19
L20
L21
L22
L23


aroc_str_pn
aroc_ent_fa
aroc_sta_au
aroc_bac_su
aroc_e_coli
aroc_hae_in
aroc_nei_me
aroc_pse_ae
β1 β2 β3 α1 β4 β5 β6 β7 α2
α3 α4 β8 α5 α6 β9 α7
β10 β11 β12 β13 β14 β15 α8 α9

against the dimer partner. Although there are many hydrogen-bonding interactions, there is just one ionic interaction at the interface, where Lys 238 from one monomer makes a water-mediated contact with the FMN phosphate group of the other.

The major component of the tetramerization interface is β sheet 1 from each monomer, which interacts with the equivalent portion of an adjacent dimer to form a β sandwich structure (Figure 1B). Several other features are also involved in the dimer-dimer interaction, most notably extended loop L8, helices α2 and α3, the short β sheet 3 (formed by strands β3, β5, and β6), and residues from the loops L2, L3, L5, L6, and L7. Although much of this interface is hydrophobic, there are several significant hydrogen-bonding interactions, and two salt-bridges between Arg 13 and Asp 75 and their 2-fold related equivalents. There are further ion-pair interactions between Arg 63 and Asp 123, and Arg 120 and Glu 372.

The observed quaternary structure is consistent with the results from a series of analytical ultracentrifugation experiments on *apo* CS, in which the apparent molecular weight ranged from 120 kDa at 2 μM concentration, to 160 kDa at 14 μM (J.M., unpublished data). The raw data showed an excellent fit to a dimer-tetramer equilibrium with a K_d of 0.8 μM.

The CS active site is at the interface between β sheet 2 and one end of the internal layer of α helices. At this point, close to the internal pseudo 2-fold axis, the two central helices diverge to leave a small hydrophobic pocket, which holds the *ortho*-xylyl end of the FMN isoalloxazine ring system. The remaining interactions with FMN and EPSP are predominantly hydrophilic, and are formed by β sheet 2 and loops L1, L4, L8, L14, L20, and L22, which lack defined secondary structure. FMN and EPSP are closely associated with each other and a considerable degree of the binding surface of each ligand is in contact with the other.

Two further FMN molecules were located in a shallow groove formed at the surface of one of the CS dimers (Labeled *1* in Figure 1B). The isoalloxazine rings of both flavins are buried at the base of the groove, while the phosphate groups are involved in crystal contacts with a neighboring tetramer. Residues on the surface of the groove that interact with FMN, notably Tyr 266, are not conserved in other species, which suggests that this interaction is not significant and simply represents a fortuitous crystal contact.

FMN Binding

Experiments with *E. coli* CS have shown that oxidized FMN has a K_d of around 30 μM, which decreases to around 20 nM in the presence of EPSP (Macheroux et al., 1998). The CS crystal structure has revealed precisely why these two values are so dissimilar, and is also consistent with observations that the binding of FMN and EPSP is ordered (Macheroux et al., 1996b). The ribityl-phosphate portion of FMN is buried deeply into the enzyme, while the entire *re* face and much of the ortho-xylyl ring of FMN contact the surface of the FMN binding site (Figure 3A). Only a part of the *si* face of the isoalloxazine ring system, corresponding to an area of 51 $Å^2$ or 7% of the surface of the FMN molecule, is therefore exposed when FMN is bound, and it is this portion of FMN that makes contact with EPSP. The binding of EPSP and the correlated tightening of the active site prevent solvent access to almost the entire FMN molecule. Just 3 $Å^2$, or less than 1% of the FMN molecular surface, is accessible once EPSP is bound. Clearly the binding of EPSP blocks any possible exit of FMN from the active site, and the induced structural changes in the active site allow FMN to make further interactions with CS. These factors are manifested in the decrease in the K_d of FMN, suggesting a greater than 1000-fold decrease in the off rate of FMN in the presence of EPSP.

FMN makes few specific polar interactions with the protein, the majority of which are contacts between the hydroxyl and phosphate oxygens of the ribityl chain and loops L8, L14, and L20, as shown in Figure 3A. The FMN phosphate sits at the dimerization interface and makes a number of contacts with residues from the adjacent monomer. The side chain of Lys 311 and the main chain nitrogen of Ala 252 make direct contact with the phosphate oxygens, as does the main chain nitrogen of Gly 296′ from the adjacent monomer, while the side chain of Lys 238′ interacts via a water molecule.

In addition to these contacts with the protein, there are a considerable number of solvent molecules close to both the phosphate and ribityl regions of FMN. These water molecules are discrete and well ordered, and many mediate interactions between FMN and the surrounding residues. FMN oxygens O5* and O4* do not make any direct interactions with the protein, but are coordinated by several solvent molecules. Likewise, oxygen O3* makes no interactions with SpCS, but is involved in a strong intermolecular hydrogen bond with one of the phosphate oxygens, which is likely to stabilize FMN in the conformation present in the active site. Oxygen O2* coordinates the side chain nitrogen of conserved Asn 251, and a second hydrogen bond with O11 of EPSP is the only direct interaction between substrate and cofactor.

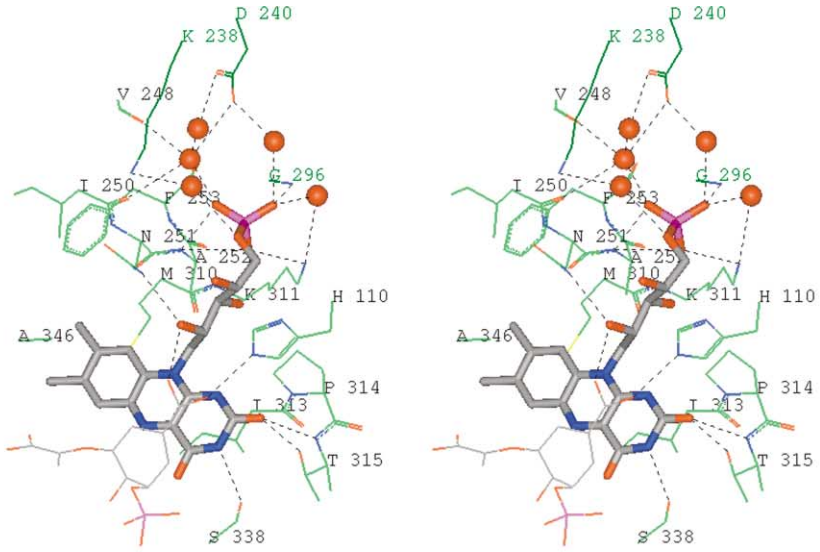
In contrast with the ribityl chain of FMN, the isoalloxazine ring system is hydrophobic and makes few specific interactions with the protein, but nevertheless it buries a considerable area of hydrophobic surface by packing

Figure 2. Topology and Sequence Alignment of Chorismate Synthase

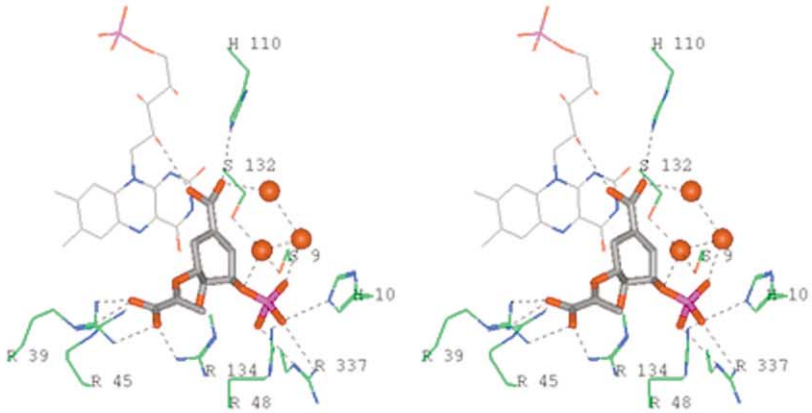
(A) Topology of CS. β strands are shown in light blue, and helices in dark blue. Strands that constitute β sheets are highlighted by red boxes. The CS core structure comprises the central column of β sheet 1, helices α1, α4, α8 and α7, and β sheet 2 in a 4β-4α-4β sandwich.

(B) Sequence conservation of CS in bacteria. The sequence alignment demonstrates the conservation within chorismate synthases from a selection of pathogenic gram-positive and gram-negative bacteria. The secondary structure of CS from *S. pneumoniae* is indicated below the alignment, and is colored in the same manner as Figure 2A. Numbering is relative to CS from *S. pneumoniae*. Active site residues that have been discussed in the text are highlighted in yellow, which demonstrates that there is considerable conservation in the CS active site. Sequence alignment produced using MULTALIN (Corpet, 1988) Residues shown in red have a correlation of at least 90% with the consensus sequence (data not shown) and are therefore highly conserved. Residues shown in blue have 50% or more correlation with the consensus sequence and are therefore moderately conserved.

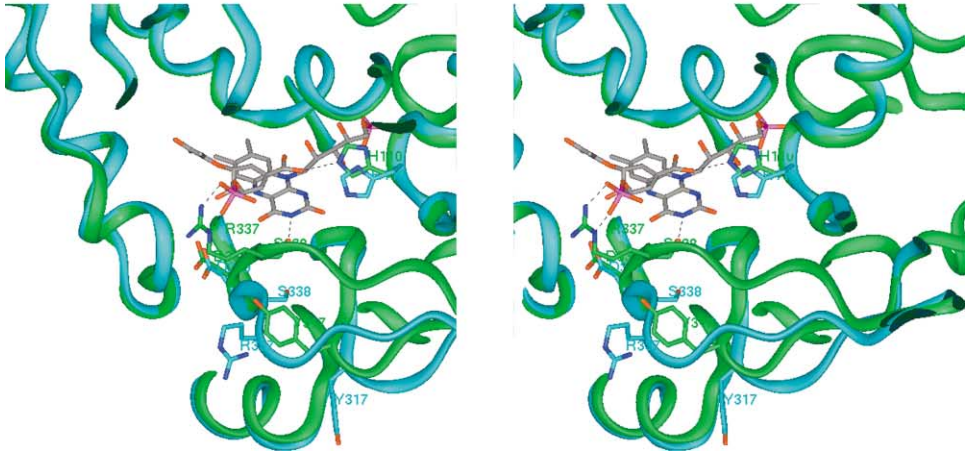
A


D 240
K 238
V 248
I 250
G 296
F 253
N 251
A 252
M 310
K 311
H 110
A 346
P 314
I 313
T 315
S 338
D 240
K 238
V 248
I 250
G 296
F 253
N 251
A 252
M 310
K 311
H 110
A 346
P 314
I 313
T 315
S 338

B


H 110
S 132
S 9
H 10
R 39
R 45
R 134
R 337
R 48
H 110
S 132
S 9
H 10
R 39
R 45
R 134
R 337
R 48

C


R337
S338
Y317
R337
S338
Y317

the *re*-face against the complementary surface provided by one end of β sheet 2. Unusually for a flavin binding protein, there are no pi-stacking interactions between protein and the flavin rings. Instead small hydrophobic residues Ala 342 (data not shown), Ala 346 (data not shown), Ala 345, Ala 252, Ile 313, and Met 310 form the binding surface (Figure 3A). This could provide additional flexibility necessary for the protein to accommodate reduced FMN, in which the isoalloxazine ring system may be bent into a "butterfly" conformation. The lack of specific contacts between CS and the FMN ribityl chain would also allow some flexing of the FMN molecule. Although butterfly flavin conformations have been observed in a number of crystal structures under reducing conditions (Lennon et al., 1999), there are also examples of reduced flavins that are almost planar (Kyte, 1995). In addition, it has been demonstrated that planar oxidized and highly puckered reduced flavins can be accommodated in the same active site without the need for changes in protein conformation (Lennon et al., 1999).

The N1-O2 locus is close to the side chain of the completely conserved His 110. These atoms are generally involved in the exchange of protons between flavins and enzymes, and hence in changes to the electronic state of flavins (Edmondson and Ghisla, 1999), which suggests a catalytic role for His 110. It is clear from the open structure that it is O2, and not N1, which forms a hydrogen bond with His 110. Differences between open and closed active sites demonstrate that the side chain of His 110 has some conformational flexibility, and interacts with FMN in the open form but with O12 of EPSP in the closed form (As shown in Figure 3C).

Superimposition of the open and closed active sites demonstrates that the binding of EPSP is associated with conformational changes in loops L20 and L22 (Figure 3C). Loop L20 bears Thr 315, whose side chain makes a hydrogen bond to O2 of FMN in the closed active site. In the open form, Thr 315 has a different rotamer, which places the hydroxyl some distance from FMN, and is instead interacting with the side chain of Ser 338 on loop L22. Residue 315 is conserved as either threonine or serine in all species (Figure 2B), and is adjacent to the completely conserved Arg 107 (Figure 3A). This pair of residues interacts with a number of conserved water molecules as well as O2 of FMN, and may therefore be catalytically important (See Mechanistic Implications).

The changes to loops L20 and L22 are clearly correlated, although the greatest positional changes occur for residues between Tyr 331 and Pro 340 on L22. Tyr 317 is the only residue on L20 that shows a significant conformational change (Figure 3C). In the open active site this residue has a solvent-exposed side chain, but a change of rotamer in the closed form allows the Tyr 317 side chain to pack against the new conformation of loop L22 and possibly helps to maintain the closed form. Loops L20 and L22 are highly conserved (Figure 2B), and most of the interacting residues from both loops are conserved in all species. Arg 337 and Asp 339 are completely conserved and Thr 315 is always threonine or serine, although there are some small differences between gram-positive and gram-negative bacteria. Tyr 317 is conserved only in gram-positive species, while Ser 338 is replaced by a conserved histidine in all gram-negative bacteria and fungi. The proximity of residue 338 to FMN in the closed form of the active site suggests that this residue may play a different, and potentially important, catalytic role in those species.

In the open form, the side chains of Ser 338 and Asp 339 from L22, as well as that of Arg 45 from loop L4, are involved in water-mediated interactions with N3 and O4 of FMN. In the closed form, loop L22 moves closer to FMN, displacing the two water molecules bound to N3 and O4, and instead forming direct interactions with those atoms (Figure 3C). Asp 339 is displaced by 1.7 Å and forms a new hydrogen bond with O4 of FMN. There is a more pronounced shift of almost 3 Å in the position of Ser 338, allowing a hydrogen bond from the side chain hydroxyl to N3 of FMN.

N5 of FMN appears not to form any interactions with protein or solvent, but does sit directly under C6 of EPSP, from which the 6-*pro*-R hydrogen atom is abstracted during the reaction. The probable role of N5 is discussed in Mechanistic Implications.

EPSP Binding

In contrast with FMN, there are extensive polar interactions between EPSP and the enzyme, but few hydrophobic contacts. The EPSP site is very hydrophilic, and must present an extremely basic environment to EPSP, with six arginine and two histidine residues concentrated in a small, tightly enclosed binding site. These basic groups are clustered at the three corners of the EPSP site, reflecting the pseudo-3-fold shape and charge distribution of EPSP itself. The interactions between EPSP and the CS active site are shown in Figure 3B. EPSP has a half boat conformation, with both the

Figure 3. The Chorismate Synthase Active Site

(A) Stereo representation of interactions made by FMN in the closed form. Active site residues are shown with green carbon atoms, FMN and EPSP with gray carbon atoms. Red spheres represent conserved water positions. Hydrogen bonds are shown as dashed lines. Some residues shown are from an adjacent monomer and form interactions that may stabilize the CS dimer. These residues are shown in darker green, and are labeled with green text, while residues from the monomer, which forms the majority of the active site, have black labels.

(B) Stereo representation of interactions made by EPSP in the closed form. Active site residues are shown with green carbon atoms, FMN and EPSP with gray carbon atoms. Red spheres represent conserved water positions. Hydrogen bonds are shown as dashed lines.

(C) Stereo diagram of the overlaid Cα traces of open (cyan ribbon) and closed (green ribbon) forms, showing differences in the conformations of loops L20 and L22. EPSP and FMN are shown with gray carbon atoms. The side chains of residues His 110, Tyr 317, Arg 337, Ser 338, and Asp 339 are shown for both forms (Open form: cyan carbons, closed form: green carbons). L22 shows the most significant movement, as demonstrated by the changes in the position of Arg 337 and Ser 338. Tyr 317 on loop L20 adopts different side chain conformations in open and closed forms, and may have a role in maintaining the closed conformation of loop L22. Hydrogen bond interactions between ligands and the highlighted residues in the closed form are shown as dashed lines. Corresponding interactions in the open form are omitted for clarity.

departing C6-*pro*-R proton and the phosphate group in axial positions.

The enol-pyruvyl moiety sits in an enclosed binding pocket, and interacts primarily with three conserved arginine residues. There is a salt-bridge interaction with Arg 39, with N-O separations of 2.6 Å (NH1–O91) and 2.9 Å (NH2–O92). In addition, there are further hydrogen bonds from O91 to NH2 of Arg 45 (2.7 Å) and from O92 to NH1 of Arg 134 (3.1 Å). O5 of EPSP makes an additional interaction with NH2 of Arg 45, and the methylene sits in a pocket formed by the aliphatic portions of the Arg 134 and Arg 48 side chains.

The relative orientations of the C10 carboxylate group and the FMN isoalloxazine ring system suggest that there is a favorable stacking interaction between the conjugated π systems of the two molecules. The C10 carboxylate makes hydrogen bonds to O2* of FMN and His 110, as already described. The latter interaction is a direct hydrogen bond in the closed form but in the open form the conformation of the His 110 side chain is different and allows a direct hydrogen bond with FMN and a water-mediated contact with O12. O12 of EPSP and the side chain of conserved Arg 107 are also involved in coordinating a water molecule, which may form part of a conserved chain of solvent molecules allowing communication and proton transfer between His 110 and the EPSP phosphate.

The binding of the phosphate group is clearly determined by the conformation of loop L22, and in particular the residues between Tyr 331 and Pro 340. The guanidinium group of the Arg 337 side chain makes a salt-bridge interaction with phosphate oxygens O2P and O3P when in the closed conformation. Arg 337 sits at the apex of the mobile portion of L22, and the interacting atoms are displaced by almost 10 Å out of the active site in the open form (Figure 3C). These two oxygen atoms make additional contacts with the side chains of Arg 48 and His 10 in both the open and closed forms of the enzyme. In the open form, the remaining phosphate oxygen, O1P, makes no direct interactions with CS, and is surrounded by solvent molecules. However, in the closed form, there is a hydrogen bond between O1P and the main chain carbonyl of Arg 337, which indicates that this atom must be protonated. (This interaction is omitted from Figures 3B and 3C for clarity). The EPSP phosphate must therefore carry a single-negative charge in the closed form of the active site.

In both open and closed forms of the active site, O3 of EPSP makes a hydrogen bond to another conserved water molecule, which is the last in the solvent chain linking His 110 to the phosphate. This water molecule is additionally coordinated by the side chains of the completely conserved residues Ser 9 and Ser 132 and is therefore ideally placed to perform a significant role in the overall mechanism (see Mechanistic Implications). The positions of the conserved water molecules are shown in Figure 3B.

The EPSP hydroxyl oxygen, O4, makes no direct interactions with the protein, and appears to make little contribution to binding. It forms a single water-mediated hydrogen bond with the side chain of Asp 339. More significantly, O4 is within 2.8 Å of one of the phosphate oxygen atoms, and may therefore play a role in stabilization of the phosphate leaving group and promotion of bond cleavage. Theoretical studies on analogs of EPSP predicted that an interaction of this type would be important to the CS mechanism (Dmitrenko et al., 2001).

Mechanistic Implications

CS catalyses the anti-1,4-elimination of phosphate and the C6-*pro*-R hydrogen of EPSP. No other known enzyme catalyses a reaction of this type, and a number of mechanisms have been proposed to account for the unique nature of the reaction (Bornemann et al., 1996b, Macheroux et al., 1999). In recent years, biochemical and kinetic investigations have narrowed the range of possibilities, and the reaction is now understood to involve an unstable, short-lived intermediate formed upon initial loss of phosphate from EPSP. There now appear to be just two possible identities for this intermediate, as outlined in Figure 4. Recent studies have provided evidence that this intermediate is a radical, formed by the transfer of a single electron from FMN, and therefore associated with a semiquinone FMN intermediate (Osborne et al., 2000), the scheme shown in Figure 4 route B. However, the alternative scheme, which involves a cationic intermediate as shown in Figure 4 route A, remains equally possible. The CS crystal structure does not provide sufficient evidence to discriminate between the two potential mechanisms, and although it is of the inactive oxidized FMN-EPSP complex, it identifies the residues that are critical in the overall reaction and will guide attempts to determine the mechanism conclusively.

Previous experiments on cyclohexene systems have shown that concerted 1,4-eliminations occur with predominantly *syn* stereochemistry (Hill and Bock, 1978). This evidence, in addition to a number of kinetic investigations (Bornemann, 2002), suggests the CS mechanism is highly likely to be nonconcerted. The reaction is therefore expected to be a two-stage process in which the first step is cleavage of the C3-O3 bond to release phosphate, also producing an unstable intermediate, which collapses to form chorismate in the second step by cleavage of the 6*proR*-C-H bond. The orientation of FMN in the CS structure suggests that it could provide stabilization of any electron-deficient intermediate of EPSP, consistent with either of the nonconcerted mechanistic schemes in Figure 4.

FMN has been shown to bind to *apo*-CS in the monoanionic-reduced form ($FMNH^-$) (Macheroux et al., 1996a). The binding of EPSP induces changes in protein conformation, concurrent with a protonation of FMN to give the neutral reduced form ($FMNH_2$). Comparison of the open and closed forms of the enzyme in the CS structure provides a rationale for the observed changes to FMN. It is clear that His 110 coordinates FMN O2 in the open form of the enzyme, but shifts in order to contact O5 of EPSP and the main chain nitrogen of Ala 111 in the closed form. His 110 is therefore likely to be the active site acid which protonates FMN O2 and neutralizes the reduced flavin in the first step of the reaction.

The interaction between His 110 and FMN suggests that the open form of the crystal structure provides a

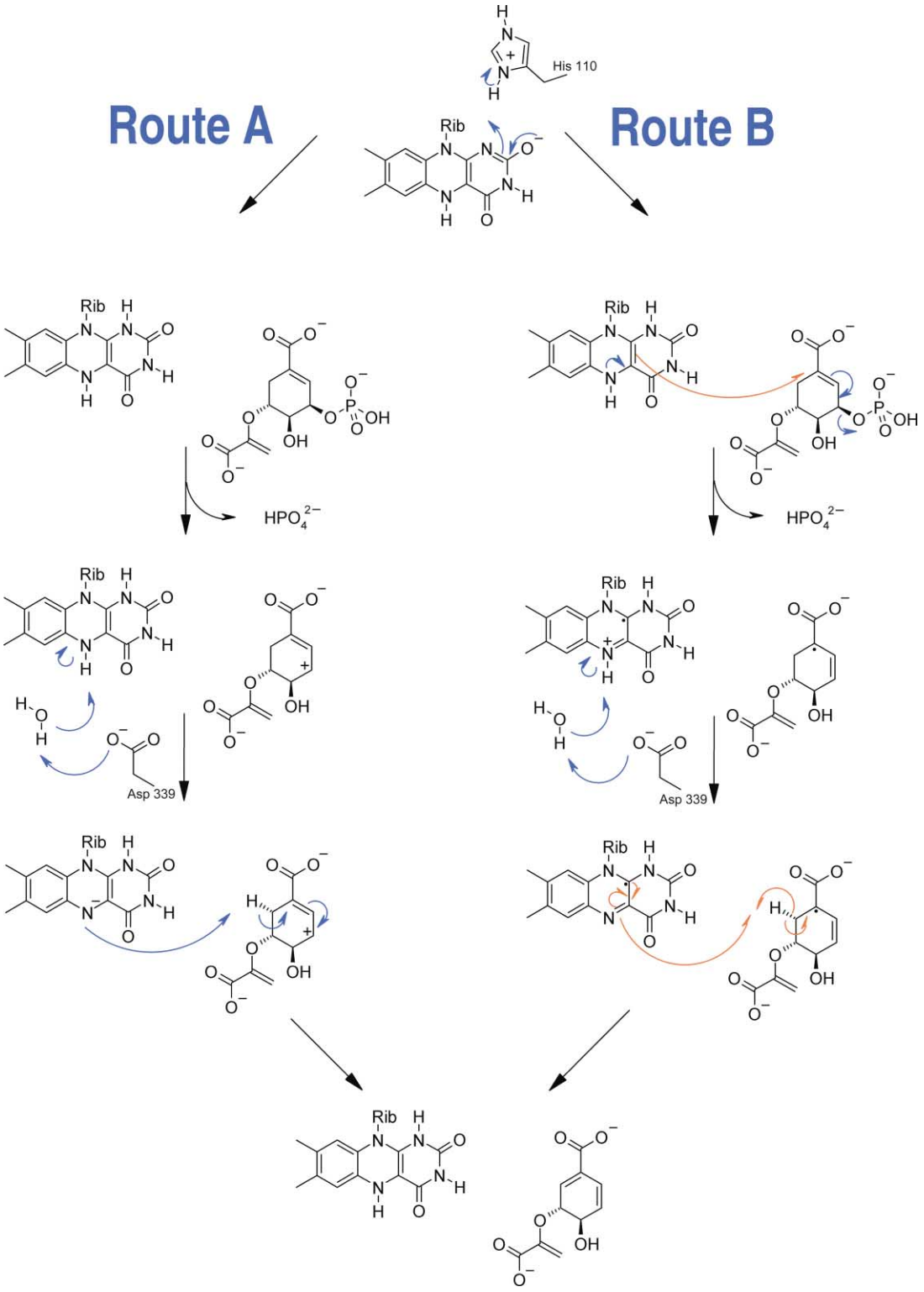


Figure 4. Mechanism of CS

Two potential reaction mechanisms are shown. Route A involves a cationic EPSP intermediate and has no requirement for radical species. This route is discussed in considerable detail in the text. Route B involves radical intermediates at both EPSP and FMN, and is also discussed in the text. In both cases residues that the crystal structure indicates are likely to be involved in the mechanism are also shown.

snapshot of the active site either before the binding of EPSP, or more probably, during the structural changes associated with the binding of EPSP. Comparison of the open and closed forms of the enzyme provides a great deal of information about which regions of the CS active site are likely to be mobile during the course of the reaction, which is of considerable bearing in the elucidation of the mechanism. Specifically, differences between open and closed active sites have demonstrated that loops L20 and L22 change conformation on binding of EPSP (Figure 3C), and that residues from these loops play important roles in the overall reaction.

Ser 338 and Asp 339 from L22 are displaced and only interact directly with FMN once EPSP is bound. Arg 337 is also displaced, and forms a number of interactions with the phosphate oxygens of EPSP, via both its main chain carbonyl and its side chain guanidyl groups, in the closed form of the active site. Each of these residues play a role in the overall CS mechanism, and Arg 337 appears to be particularly important. The EPSP phosphate is also coordinated by His 10, Arg 48, and a number of solvent molecules, whose positions are conserved in both binary (J.M., unpublished data) and ternary complexes.

Three conserved solvent molecules form a hydrogen-bonded chain which link phosphate oxygen O1P both to the carboxylate oxygen O12 of EPSP and to the side chain of His 110 (Figure 3B). The conformational change exhibited by His 110 upon EPSP binding displaces solvent, but not from these conserved positions. One of these water molecules is held in a position close to O3 of the EPSP phosphate by its interactions with the completely conserved residues Ser 9 and Ser 132. It is ideally placed to have a role in the exchange of protons between the EPSP phosphate and FMN, His 110, or possibly the C10 carboxylate of EPSP. The conserved solvent molecules may therefore be required to provide communication between FMN and EPSP via His 110, and to increase the potential for the EPSP phosphate to leave. His 10 is also likely to play a role in this process, probably by providing a proton to neutralize the charge on the departing phosphate group.

Molecular orbital calculations (Kitzing et al., submitted) have shown that the neutral reduced FMN would localize considerable electron density around C4a and C10. As the separation of C4a of FMN and C1 of EPSP is around 3.5 Å, increased electron density around this atom would be consistent with either mechanism shown in Figure 4. The build-up of positive charge at the C1-C3 locus associated with the loss of phosphate in the cationic mechanism (Figure 4, route A) would be further stabilized by this electron density, while the proximity of C4a (FMN) to C1 (EPSP) would facilitate the single electron transfer required to initiate loss of phosphate in the radical mechanism (Figure 4, route B).

The second step in the reaction mechanism is the abstraction of the C6-*pro*-R hydrogen as a proton or a hydrogen atom, producing chorismate and neutral-reduced FMN. The ground state binary complex is then regenerated by deprotonation of FMN once more. The departure of the cleaved phosphate moiety necessitates conformational changes in L20 and L22, and a reopening of the active site. However, the lifetime of the reactive intermediate is likely to be so short that these conformational changes release both phosphate and chorismate products, consistent with kinetic observations (Bornemann et al., 1996a).

Although the intermediates in either of the possible reaction schemes (Figure 4) would be stabilized by conjugation, each of the C4 and C6 protons would be acidic. The π system of either allylic intermediate would have optimal overlap with the axial bonding orbitals of the C6-*pro*-R and C4 protons, which would therefore be more acidic than the equatorial C6-*pro*-S proton. The C4 position is remote from any active site bases, which is consistent with the specificity of CS for deprotonation at C6. N5 of FMN is positioned directly below the C6-*pro*-R hydrogen of EPSP and is ideally positioned to remove this proton from the transient intermediate. The distance between N5 of FMN and C6 of EPSP is around 3.5 Å for oxidized FMN, and any change to a butterfly conformation for reduced FMN would be likely to reduce this distance.

N5 interacts with two completely conserved residues, Arg 45 and Asp 339, via another water molecule whose position is conserved in several CS structures (J.M., unpublished data). Asp 339 is ideally placed to act as a base and facilitate the removal of the acidic proton from C6 of EPSP, collapsing the unstable intermediate to produce chorismate. The critical role likely to be played by N5 explains the observation that CS is inactive with the 5-deaza analog of FMN (Lauhon and Bartlett, 1994). The regeneration of $FMNH^-$ is likely to be concomitant with the reopening of the active site, probably involving deprotonation by His 110.

The FMN semiquinone intermediate required by the radical mechanism (Figure 4B) would be highly acidic and deprotonation by Asp 339 would therefore occur readily. In contrast, the pK_a of the N5 proton in the cationic mechanism (Figure 4A) would be expected to be higher and deprotonation would be more difficult. However, changes to the environment of FMN could reduce the pK_a and make the deprotonation step possible. Closure of the active site and the associated motion of His 110 away from FMN prevent deprotonation at the other possible location, the N1-O2 locus. In addition, overlap of the cationic EPSP intermediate with the FMN isoalloxazine ring system would also be likely to reduce the pK_a of the N5 proton.

The alteration of the central ring from cyclohexene to cyclohexadiene clearly affects the shape of the intermediate, which in turn must result in a change in affinity. Despite cleavage of phosphate and possibly partial reopening of the active site, the intermediate remains bound while the proton is abstracted. In contrast, it has proved impossible to generate a CS product structure in cocrystallization and soaking experiments, which is consistent with a much lower affinity for chorismate than for EPSP, as previously reported (Macheroux et al., 1996b). This suggests that a significant contribution to the binding of both EPSP and the reaction intermediate would come from shape complementarity between their aplanar central rings and residues Ser 132 and Ser 9, which primarily form the surface against which they pack.

It is clear from a combination of the CS structure and

the level of sequence conservation that the CS active site revealed in this structure is likely to be highly similar to those of other CS enzymes. Alignment of CS sequences from a range of pathogenic bacteria (Figure 2B) demonstrates that conservation across species is generally good. The two histidine residues present in the active site, His 10 and His 110, are conserved across all known species, and His 110 forms part of a characteristic CS signature sequence (Macheroux et al., 1999). In addition all but one of the arginine residues present in the active site are also conserved across bacterial species. These include Arg 39, Arg 45, and Arg 134, which coordinate the enol-pyruvate EPSP moiety, Arg 337, which contacts the phosphate of EPSP in the closed conformation, and Arg 107, which appears to be involved in maintaining water structure around EPSP. The other residues, which form hydrogen bonds to FMN in the closed form, are Asp 339, which is invariant, and Thr 315, which is conserved as threonine or serine in all CS sequences.

The two residues, which form hydrogen bonds with EPSP or FMN, but are not invariant, are Arg 48, which contacts the EPSP phosphate, and Ser 338, which forms a hydrogen bond with N3 of FMN. Arg 48 is conserved in all gram-positive bacteria, but is always serine or threonine in other species. Interestingly, an arginine residue is conserved at position 50 in these other species, and it is possible that it performs the function of Arg 48 in those species. Similarly, Ser 338 is conserved in gram-positive bacteria but is replaced by an invariant histidine in other species. A histidine in this position would be close to N3 of FMN, O12 of EPSP, and to the side chain of Arg 107.

The importance of the two active site histidine residues has been addressed by site-directed mutagenesis (Kitzing et al, submitted). This additional evidence is consistent with both histidine residues playing critical mechanistic roles, and also demonstrates that His 110 is involved in FMN binding while His 10 is important for EPSP binding.

We have obtained further evidence for the importance of His 110 and His 10 to CS activity from chemical modification experiments. Four residues susceptible to chemical modification were tested, in the absence of substrate and cofactor, to assess their role in catalysis (S.A., unpublished data) using standard reagents (Lundblad, 1996). These studies revealed that modification of histidine residues with diethyl pyrocarbonate leads to rapid loss of all (>95%) activity. Modification of tyrosine with tetranitromethane shows a less rapid loss of total activity and modification of arginine with phenylglyoxal leads to only a moderate loss of activity (50%). Modification of lysine residues with *N*-hydroxysuccinamide had no effect (data not shown). These studies support the role of histidine in catalysis, and the results observed from modification of tyrosine may reflect the critical role played by Tyr 317 in maintaining the closed form of the enzyme.

Although bacterial chorismate synthases are monofunctional, requiring reduced FMN for activity, the fungal enzymes have an intrinsic NADPH-dependent FMN reductase activity. It has recently been shown (Kitzing et al., 2001) that the NADPH binding site in these enzymes overlaps with that of EPSP. While a structure from a bifunctional CS will be required to determine precisely how NADPH binds, the manner in which EPSP stacks its C10 carboxylate over C4a of the isoalloxazine ring system of FMN is reminiscent of the interaction between NADPH and FMN in flavin reductases (Pereira et al., 2001) and suggests that the nicotinamide ring of NADPH may mimic the binding of EPSP.

In summary, the structure of the complex of CS with oxidized FMN and EPSP supports both of the two proposed mechanisms detailed in Figure 4. A number of key residues have been identified and their likely roles in the reaction mechanism have been inferred. Although the current structure is of the inactive oxidized form of FMN, the protein structure in the region of the active site suggests that reduced FMN could easily be accommodated without structural change. While the CS structure is unable to confirm the nature of the intermediate transition state, and hence the fine detail of any electron transfer steps in the mechanism, identification of the relative orientations of FMN, EPSP, and the surrounding protein will greatly facilitate the eventual elucidation of the CS mechanism. The structure is being used as the basis for a rational drug design program, in order to identify novel inhibitors with antibacterial properties. Comparison with CS structures from a range of bacterial species (J.M., unpublished data) demonstrate that the active site is very highly conserved and thus CS is potentially an excellent broad spectrum drug target.

Chorismate synthase is an important and fascinating enzyme for several reasons: its unusual catalytic activity; its critical role in metabolism; and its potential as a broad-spectrum target for anti-infective drugs. The elucidation of its structure at high resolution with bound substrates provides the first detailed insight into how it may function and provides the basis for developing truly novel therapeutic entities for infectious diseases.

Experimental Procedures

The aroC gene of *S. pneumoniae* was identified based on its homology to other known CS genes and proteins from nonannotated genomic sequences of *S. pneumoniae* deposited in the public databases. The gene was cloned into expression vector pET22b, and soluble, active protein was overproduced in the *E. coli* strain BL21 (DE3). SpCS protein was purified using a modified protocol based on that used to purify SaCS (Horsburgh et al., 1996). Enzyme activity was assayed as previously described (Webster et al., 2002 [Patent GB2374414])

Selenomethionine-incorporated and unmodified SpCS were crystallized under similar conditions. Crystals were grown by vapor diffusion with hanging drops. The protein was concentrated to 6–10 mg/ml in 10 mM Tris-HCl [pH 7.5], 20 mM KCl, 0.5 mM DTT, and 2 mM EDTA. Reservoirs containing 9% PEG 8000, 10% ethylene glycol, 100 mM HEPES [pH 7.5] were equilibrated against drops composed of a 1:1 mixture of reservoir solution and protein solution containing 2 mM FMN, 1 mM EPSP and 12 mM $(NH_3)_6CoCl_3$. Block-shaped crystals of dimensions 0.4 × 0.4 × 0.4 mm grew in 2–4 weeks. These crystals belonged to the monoclinic space group $P2_1$, with cell dimensions a = 81.1 Å, b = 124.6 Å, c = 85.2 Å, β = 115.15°, and contained one SpCS tetramer per asymmetric unit, giving a solvent content of 40%.

All data sets used to solve the CS structure were collected at ESRF, Grenoble, France, using a Mar charge-coupled detector, and were processed and reduced using programs of the HKL (Otwinowski and Minor, 1997) and CCP4 (CCP4, 1994) suites. A three-wavelength MAD (Multiwavelength Anomalous Dispersion) dataset was

collected to 2.7 Å from a single SeMet-incorporated crystal. The high-resolution dataset was collected to 1.9 Å from a second SeMet crystal. 30 of 48 selenium atom positions were identified using Shake'n'Bake v1.5 (Miller et al., 1994) and programs of the CCP4 suite were used to locate the remaining selenium atom positions, refine these atomic parameters and to generate MAD phases. Initial maps were of sufficient quality to determine matrices describing the noncrystallographic symmetry (NCS) within the crystal. A combination of solvent-flattening, histogram-matching, phase extension and 4-fold NCS averaging using DM produced traceable maps with a mean figure of merit (FOM) of 0.77 to 2.0 Å resolution. Crystallographic data used to solve the CS structure are detailed in Table 1.

The protein model was constructed using iterative cycles of model building using QUANTA (Accelrys Inc., San Diego, CA) and refinement using REFMAC (Murshudov et al., 1997). Four-fold NCS restraints were initially applied but were relaxed in the later stages of refinement, as it became apparent that there were differences between NCS-related molecules. The final model contains all 388 residues from each of the four monomers. All main chain atoms and most of the side chain atoms are well defined in electron density. Each of the four active sites contains FMN and EPSP. In addition, two other FMN molecules have been identified bound to the surface of the protein and involved in crystal contacts. The final model also contains seven ethylene glycol (ETG) molecules, nine hexaamine cobalt (III) chloride (NCO) molecules and 1929 water molecules. The R-factor of the refined model is 15.69% (R_{free} = 22.24%) and the geometry of the model has been verified using PROCHECK (Laskowski et al., 1993). The final refinement statistics can be found in Table 2.

Acknowledgments

We thank Gordon Leonard and Andy Thompson of ESRF for help with data collection and MAD solution. We thank Olwyn Byron for performing ultracentrifugation experiments. We also gratefully acknowledge John Coggins, Peter Macheroux, John Barker, and Bill Primrose for useful discussions and critical reading of the manuscript. We are indebted to referees for their insightful comments.

Received: May 23, 2003
Revised: September 9, 2003
Accepted: September 12, 2003
Published: December 2, 2003

References

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The protein data bank. Nucleic Acids Res. *28*, 235–242.

Bornemann, S., Lowe, D.J., and Thorneley, R.N.F. (1996a). The transient kinetics of *Escherichia coli* chorismate synthase: substrate consumption, product formation, phosphate dissociation, and characterization of a flavin intermediate. Biochemistry *35*, 9907–9916.

Bornemann, S., Lowe, D.J., and Thorneley, R.N.F. (1996b). *Escherichia coli* chorismate synthase. Biochem. Soc. Trans. *24*, 84–88.

Bornemann, S. (2002). Flavoenzymes that catalyse reactions with no net redox change. Nat. Prod. Rep. *19*, 761–772.

Carpenter, E.P., Hawkins, A.R., Frost, J.W., and Brown, K.A. (1998). Structure of dehydroquinate synthase reveals an active site capable of multistep catalysis. Nature *394*, 299–302.

CCP4 (Collaborative Computational Project 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D. Biol. Crystallogr. *55*, 849–861.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. *16*, 10881–10890.

Dmitrenko, O., Wood, H.B., Jr., Bach, R.D., and Ganem, B. (2001). A theoretical study of the chorismate synthase reaction. Org. Lett. *3*, 4137–4140.

Edmondson, D., and Ghisla, S. (1999) Flavoenzyme structure and function. In Flavoprotein Protocols, S.K. Chapman and G.A. Reid, eds. (Totowa, NJ: Humana press), pp. 157–179.

Gourley, D.G., Shrive, A.K., Polikarpov, I., Krell, T., Coggins, J.R., Hawkins, A.R., Isaacs, N.W., and Sawyer, L. (1999). The two types of 3-dehydroquinases have distinct structures but catalyze the same overall reaction. Nat. Struct. Biol. *6*, 521–525.

Haslam, E. (1993) Shikimic Acid: Metabolism and Metabolites. (Chichester, UK: John Wiley & Sons).

Hill, R.K., and Bock, M.G. (1978). Stereochemistry of 1,4-conjugate elimination reactions. J. Am. Chem. Soc. *100*, 637–639.

Horsburgh, M.J., Foster, T.J., Barth, P.T., and Coggins, J.R. (1996). Chorismate synthase from Staphylococcus aureus. Microbiol. *142*, 2943–2950.

Kitzing, K., Macheroux, P., and Amrhein, N. (2001). Spectroscopic and kinetic characterization of the bifunctional chorismate synthase from *Neurospora crassa*. J. Biol. Chem. *276*, 42658–42666.

Krell, T., Coggins, J.R., and Lapthorn, A.J. (1998). The three-dimensional structure of shikimate kinase. J. Mol. Biol. *278*, 983–997.

Kyte, J. (1995) Mechanism in Protein Chemistry (New York: Garland Publishing Inc.)

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Lauhon, C.T., and Bartlett, P.A. (1994). Substrate analogs as mechanistic probes for the bifunctional chorismate synthase from *Neurospora crassa*. Biochemistry *33*, 14100–14108.

Lennon, B.W., Williams, C.H., Jr., and Ludwig, M.L. (1999). Crystal structure of reduced thioredoxin reductase from *Escherichia coli*: structural flexibility in the isoalloxazine ring of the flavin adenine dinucleotide cofactor. Protein Sci. *8*, 2366–2379.

Lundblad, R. (1996). Chemical modification of amino acid side chains, In Proteins Labfax, N.C. Price, ed. (Oxford, UK: Bios Scientific Publishing), pp. 287–298.

Macheroux, P., Bornemann, S., Ghisla, S., and Thorneley, R.N.F. (1996a). Studies with flavin analogues provide evidence that a protonated reduce FMN is the substrate-induced transient intermediate in the reaction of *Escherichia coli* chorismate synthase. J. Biol. Chem. *271*, 25850–25856.

Macheroux, P., Petersen, J., Bornemann, S., Lowe, D.J., and Thorneley, R.N.F. (1996b). Binding of the oxidized, reduced and radical flavin species to chorismate synthase. An investigation by spectrophotometry, fluorimetry and electron paramagnetic resonance and electron double resonance spectroscopy. Biochemistry *35*, 1643–1652.

Macheroux, P., Schönbrunn, E., Svergun, D.I., Volkov, V.V., Koch, M.H.J., Bornemann, S., and Thorneley, R.N.F. (1998). Evidence for a major structural change in *Escherichia coli* chorismate synthase induced by flavin and substrate binding. Biochem. J. *335*, 319–327.

Macheroux, P., Schmid, J., Amrhein, N., and Schaller, A. (1999). A unique reaction in a common pathway: mechanism and function of chorismate synthase in the shikimate pathway. Planta *207*, 325–334.

Madej, T., Gibrat, J.-F., and Bryant, S.H. (1995). Threading a database of protein cores. Proteins *23*, 356–369.

Michel, G., Roszak, A.W., Sauvé, V., Maclean, J., Matte, A., Coggins, J.R., Cygler, M., and Lapthorn, A.J. (2003). Structures of shikimate dehydrogenase AroE and its paralog YdiB: a common structural framework for different activities. J. Biol. Chem. *278*, 19463–19472.

Miller, R., Gallo, S.M., Khalak, H.G., and Weeks, C.M. (1994). *SnB*: crystal structure determination via *Shake-and-Bake*. J. Appl. Crystallogr. *27*, 613–621.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D. Biol. Crystallogr. *53*, 240–255.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. (1997). CATH-A Hierarchic classification of protein domain structures. Structure *5*, 1093–1108.

Osborne, A., Thorneley, R.N.F., Abell, C., and Bornemann, S. (2000). Studies with substrate and cofactor analogues provide evidence for a radical mechanism in the chorismate synthase reaction. J. Biol. Chem. *275*, 35825–35830.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In Methods Enzymol. Volume 276: Macromolecular Crystallography Part A, C.W. Carter, Jr., and R.M. Sweet, eds. (New York: Academic Press), 307–326

Pereira, P.J.B., Macedo-Ribeiro, S., Párraga, A., Pérez-Luque, R., Cunningham, O., Darcy, K., Mantle, T.J., and Coll, M. (2001). Structure of human biliverdin IXβ reductase, an early fetal bilirubin IXβ producing enzyme. Nat. Struct. Biol. *8*, 215–220.

Roberts, F., Roberts, C.W., Johnson, J.J., Kyle, D.E., Krell, T., Coggins, J.R., Coombs, G.H., Milhous, W.K., Tzipori, S., Ferguson, D.J., et al. (1998). Evidence for the shikimate pathway in apicomplexan parasites. Nature *393*, 801–805.

Shindyalov, I.N., and Bourne, P.E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng. *11*, 739–747.

Shumilin, I., Kretsinger, R., and Bauerle, R. (1999). The three dimensional structure of the phenylalanine-sensitive 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Escherichia coli*. Structure *7*, 865–875.

Stallings, W.C., Abdel-Maguid, S.S., Lim, L.W., Shie, H.-S., Dayringer, H.E., Leimgruber, N.K., Stegeman, R.A., Anderson, K.S., Sikorski, J.A., Padgette, S.R., et al. (1991). Structure and topological symmetry of the glyphosate target 5-enolpyruvylshikimate-3-phosphate synthase: a distinctive protein fold. Proc. Natl. Acad. Sci. USA *88*, 5046–5050.

White, J.V., Stultz, C.M., and Smith, T.F. (1994). Protein classification by stochastic modeling and optimal filtering of amino-acid sequences. Math. Biosci. *119*, 35–75.

Accession Numbers

Coordinates have been deposited in the Protein Data Bank (accession code 1QX0).